DOCUMENT RESUME

ED 184 271 EC 123 420

AUTHOR Bagstad, Nannette K.
TITLE Effects of Sequentially-Based Objectives and Criterion-Referenced Tests on Academic Performance. BICAR (Binford Individualized Curriculum and Analysis Research).
INSTITUTION Binford Public School, N.Dak.
NOTE 27p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Criterion Referenced Tests; *Disabilities; *Educational Objectives; Elementary Education; Exceptional Child Research; Individualized Instruction; Program Effectiveness; *Sequential Approach
IDENTIFIERS *Binford Individualized Curriculum Analysis Project

ABSTRACT

As part of Project BICAR (Binford Individualized Curriculum and Analysis Research), the study determined the effects of sequentially based objectives and criterion referenced tests on academic performance of 42 second, fourth, and sixth graders (eight of whom tested 1 to 4 years below grade level). A review of the literature supported the premise that sequentially based objectives and criterion referenced tests, if well sequenced and measurable, would assist the staff in teaching students at their individual academic level. Following administration of criterion referenced tests, Ss progressed through all objectives with continual review of subject matter. From evidence it was concluded that there was some marked change at all levels of learning and that Project BICAR had a definite effect on academic performance. Success of Project BICAR was attributed to strengths which included teachers' awareness of individual differences and continuity in teaching skills. (SBH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED184271

BICAR

(Binford Individualized Curriculum and Analysis Research)

EFFECTS OF SEQUENTIALLY-BASED OBJECTIVES AND CRITERION-REFERENCED TESTS ON ACADEMIC PERFORMANCE

Binford Public School

Binford, North Dakota

Nannette K. Bagstad

BICAR Co-ordinator

December, 1979

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Lloyd K. Gjovik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EC123420

TABLE OF CONTENTS

ACKNOWLEDGMENTS iv

CHAPTER I. INTRODUCTION 1

CHAPTER II. REVIEW OF LITERATURE 4

CHAPTER III. PROCEDURES 10

CHAPTER IV. RESULTS 12

CHAPTER V. SUMMARY 20

REFERENCES 22

ACKNOWLEDGMENTS

I wish to express my appreciation to Lloyd Gjovik, Superintendent at Binford Public School, the Binford staff, Cloydell Viseth, and my mother, Nettie Bagstad, for their assistance and encouragement during the writing of my paper.

I also wish to thank Dr. Myrna Olson for assisting me during my graduate studies at the University of North Dakota.

CHAPTER I

INTRODUCTION

With the advent of Public Law 94-142 and the directive of the North Dakota Legislature to provide education for all handicapped children by 1980, schools have been forced to analyze their curriculum in light of content, direction, and instructional techniques.

With the challenge of these directives in mind, the administration at Binford Public School reviewed the educational records of each student in grades one through twelve. It was found that: (1) in four consecutive years, the freshmen class scored in the lowest quartile on the Science Research Associates Test (SRA), (2) test results from the special education staff indicated a high incidence of students with special needs, and (3) four students dropped out of school within a two year period of time.

The administration contacted the faculty, parents, students, special education personnel, and the Board of Education to form a curriculum study group to evaluate the existing curriculum and the quality of instruction.

Following the curriculum study group recommendations, the administration took a firm position on education when they decided that individualized instruction could provide stimulating learning experiences as well as providing the social and emotional development necessary for satisfactory life adjustment.

With the individualized instruction goal uppermost in their minds, the administration wrote a proposal for a three year grant through the Elementary and Secondary Education Act (ESEA) Title 4-C.

Basically, it was determined that Binford Public School could meet the state and federal mandates to provide quality education locally for all students through the approval of the ESEA Title 4-C grant, which would enable Binford to develop sequentially-based objectives and criterion-referenced tests to provide individualized instruction. Project BICAR (Binford Individualized Curriculum and Analysis Research) was approved and consequently, during July and August, 1977, the task of revising the reading, math, and language arts curriculum for grades K-8 was begun. Subsequent work was completed in all subject areas K-12 and revisions have been an initial part of the entire project during the years from 1977-1979.

While formulating sequentially-based objectives and criterion-referenced tests, the staff focused attention on the following problem areas: (1) meeting student needs, (2) providing continuity in teaching skills (and terminology used to teach those skills) from one grade to the next, (3) raising test results both locally and nationally, (4) alleviating student apathy, and (5) lowering the drop out ratio (present rate = approximately two percent of the total school population).

It was Project BICAR's goal to enable each student to progress at his own rate as he mastered levels indicated by evaluative instruments. Now, after two years of implementation of the competency-based curriculum, the writer of this study hoped to determine the success of that project. This study will determine the effects of sequentially-based objectives and criterion-referenced tests on academic performance.

BICAR Working Definitions

Following are the definitions under which Project BICAR was implemented:

(1) sequentially-based objectives: goals--without bias or prejudice--which form a regular pattern of parts.

EXAMPLE: When shown the letters--c,t, and a--, the child will
(a.) identify the letter by name
(b.) say the appropriate sound for each letter with 100% accuracy
(c.) decode the word "cat" and pronounce it correctly.

(2) criterion-referenced tests: a set of questions, problems, or exercises which determine a person's knowledge and abilities and to which a measurable judgment can be applied.

Cattell (1944) defines criterion-referenced tests as a measure that yields a score that is directly interpretable without reference to an individual's relative standing in a reference group.

EXAMPLE: Given the pre-primer list of Dolch words, the student will read the list with 95% accuracy in one minute.

(3) norm-referenced tests: a standard of achievement as represented by the median or average achievement of a large group.

CHAPTER II

REVIEW OF LITERATURE

While reviewing the literature, the writer attempted to locate material dealing with the following issues: (1) formulation of criteria for writing objectives in sequential order, (2) evaluation of performance through the use of criterion-referenced tests, and (3) implementation of individualized instruction by other schools in the nation. The preceding issues were chosen to determine whether or not the behavioral objectives and performance objectives developed by Project BICAR were comparable to similar studies pursued by other schools.

Writing Objectives

The literature varied in terms of the exact criteria to include for sequentially-based objectives and criterion-referenced tests. Most researchers contended that one should design performance objectives with an order of complexity which relate to two dimensions--content area and level of achievement (Boozer 1971). The student would progress through the curriculum by starting with the least complex area and by progressing to the most complex. The student would master all objectives in a given level prior to advancing to the next higher level.

Accordingly, in order to investigate the appropriate sequence of objectives within the curriculum in terms of learner's performance, it would be necessary to work within a setting where the units of the curriculum have been defined in "measurable behavior" (Boozer 1971).

To further support measurable behavior, Skager (1978) indicated that behavioral objectives should state the conditions confronting the examinee and the observable behavior in view of the correct response to the conditions. Skager (1978) also contended that inclusion of an arbitrary definition of mastery, which was strongly recommended by some researchers, would be inappropriate. The mastery element would confuse the specification of the content domain with the setting of an interpretive standard. Behavioral objectives leave much latitude up to the item writer (Skager 1978).

According to the Florida Department of Education, the acceptable level of performance on objectives must be determined by the schools. Once the schools have set acceptable limits for achievement, they must analyze objectives and criterion-referenced test results to determine: (1) strengths and weaknesses, (2) which objectives receive priorities on the local level, and (3) which objectives need revisions and in what order they should be considered ("A Guide to 1976-77 Statewide Assessment Results" 1976-77).

Evaluating Performance

While considering performance objectives, one must also consider evaluative measures since criterion-referenced tests provide information in terms of specific behavior mastered without reference to the performance of other pupils (Cox 1966).

Martuza (1977) contended that the first step in implementing criterion-referenced measurement was to develop, prior to instruction, a list of objectives which would identify the performance, skills, and products the instruction was designed to help students attain. This list would become the standard for judging learning success; it would be the criterion against which student's performance would be compared to judge learning accuracy.

While criterion-referenced tests were recommended by the researchers, mastery scores and test validity were open to discussion and varying interpretations. The most widely used mastery score tended to be 85% on a pre-test. If the student scored 85% on the pre-test, he skipped that unit and would continue to the next domain (Boozer 1971).

Other data indicated that 85% was a low criteria since it was within the "average range". Boozer (1971) argued, however, that only the most important objectives would be tested on a criterion-referenced test.

In the Colorado Education Accountability Act, better results were reported in the elementary grades than in the secondary grades. Reading and math scores were about the same. Overall, the Colorado Accountability Act reported that 63% of the districts' median scores were at or above the national norms (Hennes 1977).

When determining success or failure via criterion-referenced tests, it was warned by the Florida Department of Public Instruction (1976-77) that what constituted a "statistically significant" difference was not usually a satisfactory approach. When determining achievement percentages, one should look for differences of about 10-15%. In a classroom of 25-30 students, one child would represent three to four percent of the class total. Thus, if one child missed only one objective, he would score 96%; two missed objectives would merit a 92%. One can readily see the large difference one student's scores could make on the summary score.

A number of writers have concluded that content validity for criterion-referenced tests follows automatically from the rationale establishing the test domain and that no further consideration of their validity would be necessary. Thus, when considering validity, it would be wise to remember that criterion measures are more valid when they adequately represent that criterion (Crambert 1977).

Implementing Individualized Instruction

The literature showed that schools have incorporated sequenced objectives and criterion-referenced tests into their curriculum. One study outlined the necessity of setting minimal objectives for achieving basic skills in language arts and math. The other study focused on the school district's responsibility for improving education and for providing data to indicate improved pupil performance.

The Florida Department of Education stated that one of the priorities of the Florida educational system was helping all students acquire the basic skills necessary to perform the reading, writing, and math tasks required for everyday living.

Minimal objectives for math and communication skills were formulated. The catagories in each subject area formed a general heirarchy. Vocabulary skills were pre-requisites to comprehension skills; comprehension skills were pre-requisites to critical skills. Within a catagory of objectives, the objectives were presented in a progression of difficulty.

Two sets of objectives were written--Milestone Objectives, which identified the essential terminal skills to perform everyday tasks and Grade Level Objectives, which defined each milestone objective by specifying which aspect of the skill should be acquired by selected grades (three, five, eight, and eleven). Entry level performance for each grade being tested in statewide assessment were determined when the objectives were formulated.

While the same set of objectives were used for statewide assessment from 1972-76, objectives were revised and reduced. During the 1976-77 school year, new objectives were adopted by the Department of Instruction since some of the previous objectives did not meet the basic skill's definition, were inappropriate for assessment of a given grade,

or were not clearly written. The Florida Education Department determined criteria for objective achievement by developing the following rationale:

For objectives with: 1-4 items -- 100%
5-6 items -- 1 error
7-8 items -- 2 errors
larger sets -- 85%

A two day test to assess the educational program was conducted in grades three, five, eight, and eleven. The test results served to identify the immediate student needs and possible curriculum modification.

Basically, the Florida Education Department's "A Guide to 1976-77 Statewide Assessment" outlined the minimal objectives, acceptable achievement levels, and interpretation of test results to assist the schools in utilizing the results to assure successful school improvement.

In 1971, the Colorado General Assembly enacted the Education Accountability Act, which required each school district to plan and implement school improvement. Since enactment of the Accountability Act, Colorado schools have been instructed to evaluate and report pupil performance in math and reading. The educational accountability focused on local control of education since needs and priorities were best identified on the local level.

Criterion-referenced tests and/or norm-referenced tests were chosen by each individual school district to determine pupil performance. Some schools used only criterion tests or normed tests; others used a combination of tests. Seventy three percent of the school districts reported the median achievement scores at or above the national norms. Better results were reported in the elementary grades than in the secondary grades. Reading and math scores were about the same.

Summary

Basically, the literature tended to support the premises under which Project BICAR was created. Sequentially-based objectives and criterion-referenced tests, if well sequenced and measurable, would indeed assist the staff in teaching students at their individual academic level.

CHAPTER III

PROCEDURES

Grades two, four, and six at Binford Public School were chosen to participate in the study to determine the effects of sequentially-based objectives and criterion-referenced tests on academic performance.

The total number of children participating was 42. Thirty four were tested "at level" (at their appropriate level) and eight were tested "out of level" (one to four years below grade level, depending on their individual academic placement). Statistical data reported in Chapter IV does not represent the children tested "out of level" since it was the goal of the researcher to determine the academic achievement of the children who were performing within the average range.

Criterion-referenced tests were administered to students on a regular basis when the instructors determined that an appropriate amount of material had been covered. Goals for academic achievement on the criterion tests were set at 85-100% mastery, 65-85% for partial mastery, and below 65% for non-mastery. The mastery goal was set at a minimum of 85%, since the staff felt that it would allow for further learning and review of specific skills.

Project BICAR deviated from the normal procedure of using a pre-test to determine whether or not a student must work on a given objective. After a student's academic level had been determined, he progressed through all of the objectives. Pre-tests at those levels were not administered since the staff supported the belief that "overlearning" was an essential

part of the educational process. After the domain objectives were learned and/or reviewed, the criterion-referenced tests was administered. If the student showed mastery, he moved on to the next skill.

It is essential to note again that the staff maintained that frequent review of subject matter was necessary for efficient retention of knowledge. The reading and math texts used by the Binford Public School allowed for continuous learning and review of basic objectives. Thus, even when a student scored at the mastery level, overlearning became an initial part of the process.

To further assist the staff in determining academic performance, the California Achievement Tests (CAT) was given as a pre/post test. The CAT was chosen as the appropriate test to measure progress. It was found that specific subtests and measurable skills in the CAT were comparable to the objectives developed for Project BICAR. Math and reading subtests and skills measured by CAT and Project BICAR included:

(1) Phonetic Analysis and Structural Analysis - decoding,

(2) Reading Vocabulary - word catagories and meanings (same, opposite, and multi-meaning),

(3) Reading Comprehension - literal, interpretive, and critical,

(4) Math Computation - all processes (add, subtract, multiply, and divide) for whole numbers, fractions, mixed numbers, decimals, and algebraic expressions, and

(5) Math Concepts and Applications - emphasis on concepts rather than isolated facts.

An analysis of statistical data is included in Chapter IV. Specific reading and math results for grades two, four, and six will be reported. These scores are based on the pre/post tests on the California Achievement Tests which were administered during the 1978-79 school year.

CHAPTER IV

RESULTS

An analysis of variance between the pre/post test scores for the California Achievement Tests was conducted by an individual outside of the school district. Specific data in relationship to (1) pre-test scores versus post-test scores, (2) anticipated post-test scores versus actual post-test scores, and (3) national median scores versus the Binford Public School median scores further support the statistical findings.

Pre-test scores versus post-test scores (see table 1) showed an overall increase in scores for the majority of students in grades two, four, and six. The T-values (see table 2) showed statistical significance (.001 level) for all grades except grade six.

The anticipated post-test scores (see table 3) were determined to check student performance on the actual post test. On the average, grades two, four, and six scored at or above the anticipated post-test score when compared with the actual results. The anticipated scores were determined by assigning 1/3 of a standard deviation to each individual pre-test score. The standard deviation for each grade was determined by using the following formula:

$$\sqrt{\frac{\Sigma (X - \bar{X})^2}{df}}$$

X = pre-test score

$\bar{X}$ = test mean

df = degrees of freedom

A comparison of the national median versus Binford Public School median further indicated that BICAR was successful, since the average school scores were higher than the national in all grades studied except for second grade math which was only eight points below the national median. Table five outlines the median scores for both the pre/post tests.

Grade two showed the most significant increases for all areas tested (see table 2). Factors which could account for the significant scores follow: (1) firm academic groundwork was laid during the first grade, (2) greater socialization through exposure to school situations, which is extremely valuable in very rural areas, and (3) team teaching was used to a great extent in the areas of reading and math. The team consisted of the classroom teacher, Basic Skills teacher, and a teacher's aide.

Grade four showed the greatest change in math, but also showed changes of statistical significance in reading. There seemed to be no apparent reason for the math results, but the writer hypothesized that student performance could be attributed to an intensive study of basic math concepts and applications during the previous school year.

Grade six showed a significant change in language arts only. In grade six, two factors could account for failure to score significantly higher -- (1) one student should have been tested "out of level" as further testing in the fall of 1979 indicated that the academic level was too difficult for the student, and (2) when determining T-values, one must decrease the total sample by two (n-2); when this was done, the total sample was only three students. Thus, the T-values must be larger to show statistical significance.

Evidence within the statistical report submitted to the Department of Public Instruction and Binford Public School concluded that at all levels of learning, there was some marked change. Consequently, Project BICAR has had a definite effect on academic performance at Binford Public School.

TABLE 1

PRE-TEST SCORES VERSUS POST-TEST SCORES

	Reading			Math		
	X	Y	Y - X	X	Y	Y - X
Grade 2	341	363	+22	306	397	+91
	350	403	+53	306	349	+43
	346	411	+65	341	386	+45
	315	359	+44	309	343	+34
	335	368	+33	349	372	+23
	350	381	+31	301	369	+68
	345	391	+46	360	372	+12
	363	444	+81	347	397	+50
	381	411	+30	347	391	+44
	403	432	+29	325	369	+44
	309	357	+48	---	---	---
	359	411	+52	337	378	+41
	366	381	+15	325	372	+47
	385	432	+47	337	372	+35
	348	391	+43	333	354	+21
	343	444	+101	316	353	+37
	337	385	+48	309	364	+55
	403	432	+29	325	397	+72
	361	444	+83	320	356	+36
Grade 4	419	446	+27	422	455	+33
	466	488	+22	431	455	+24
	472	543	+71	424	475	+51
	414	494	+80	424	460	+36
	417	440	+23	398	396	- 2
	414	466	+52	460	516	+56
	543	558	+15	442	494	+52
	543	558	+15	452	458	+ 6
	479	577	+98	428	450	+22
	437	463	+26	426	455	+29
Grade 6	517	597	+80	530	554	+24
	521	557	+36	482	526	+44
	571	571	0	533	543	+10
	466	496	+30	418	479	+61
	571	633	+62	519	543	+24

X = pre score
Y = post score

TABLE 2

ANALYSIS OF VARIANCE

T-Values

Grade	Reading	Math	Sampling
K-12	3.61*	3.48*	103
2	9.33*	10.60*	19
4	4.49*	5.04*	10
6	3.02	3.65	5

* Statistical Significance (.001)

TABLE 3

ANTICIPATED SCORES VERSUS POST-TEST SCORES

	Reading			Math		
	Y	Yc	Y - Yc	Y	Yc	Y - Yc
Grade 2	363	378	-15	397	312	+85
	411	383	+28	386	347	+39
	403	387	+16	349	312	+37
	359	352	+ 7	343	315	+28
	368	372	- 4	372	355	+17
	381	387	- 6	369	307	+62
	391	382	+ 9	372	366	+ 6
	444	400	+44	397	353	+44
	411	418	- 7	391	353	+38
	432	440	- 8	369	331	+38
	357	346	+11	---	---	---
	411	396	+15	378	343	+35
	381	403	-22	372	331	+41
	432	422	+10	372	343	+29
	391	385	+ 6	354	339	+15
	444	380	+64	353	322	+31
	385	374	+11	364	315	+49
	432	440	- 8	397	331	+66
	444	398	+46	356	326	+30
Grade 4	446	441	+ 5	455	428	+27
	488	488	0	455	437	+18
	543	494	+49	475	430	+45
	494	436	+58	460	430	+30
	440	439	+ 1	396	404	- 8
	466	436	+30	516	466	+50
	558	565	- 7	494	448	+46
	558	565	- 7	458	458	0
	577	501	+76	450	434	+16
	463	459	+ 4	455	434	+21
Grade 6	597	532	+65	554	546	+ 8
	557	536	+21	526	498	+28
	571	586	-15	543	549	- 6
	496	481	+15	479	434	+45
	633	586	+47	543	535	+ 8

Y = post score
Yc = anticipated score

TABLE 4

ANALYSIS OF VARIANCE

T-Values

Grade	Reading	Math	Sampling
2	2.09	8.63*	19
4	2.69	4.01*	10
6	1.93	1.86	5

*Statistical Significance (.001)

TABLE 5

NATIONAL MEDIAN VERSUS BINFORD MEDIAN

		Reading			Math		
		National	BPS	Difference	National	BPS	Difference
Grade 2	Pre:	336	350	+14	333	325	- 8
	Post:	366	415	+44	358	372	+14
Grade 4	Pre:	425	466	+41	410	428	+18
	Post:	446	494	+48	434	458	+24
Grade 6	Pre:	486	521	+35	475	519	+44
	Post:	503	571	+68	496	543	+47

CHAPTER V

SUMMARY

Current Conclusions

Based on local and national evaluative data, it appears that Project BICAR has definitely been a success since the curriculum was revised to include the use of sequentially-based objectives and criterion-referenced tests.

Teaching classes through the use of objectives so that all objectives are learned and continually reviewed has been a positive factor in improved educational performance. Criterion-referenced tests have assisted the teachers in determining individual needs for continued instruction and/or a general review of basic concepts.

The pre/post norm-referenced tests have supported the local data by showing that student performance has either equalled or exceeded the median scores which were normed by school districts participating in the California Achievement Test standardization.

The success of Project BICAR might be attributed to several factors. Among them, the writer would propose the following strengths of the project:

(1) it allowed teachers to work together to create a better curriculum,

(2) it made teachers aware of individual differences and caused them to do something to make learning more meaningful,

(3) it allowed for continuity in teaching skills, and

(4) it allowed for students, faculty, administration, and parents to take an honest look at the educational system and decide that a small community can take pride in their school and their children's academic life.

Project BICAR's future is dependent upon several factors:

(1) continued implementation of sequentially-based objectives and criterion-referenced tests, even though changes in administration and staff may occur,

(2) continued revision and refinement of the objectives and tests to assure proper curriculum revitalization, and

(3) continued use of a record keeping system which allows personnel to determine individual pupil progress at a glance.

Future Studies

An extension of Project BICAR could be the addition of a computer to store the sequentially-based objectives for all areas of study and for all grade levels. The ultimate goal would be to develop computerized individual educational plans to meet the needs of all students enrolled at Binford Public School.

Much interest has been expressed by administrators in other schools. Consequently, it would be advantageous to disseminate information relative to initial planning, procedures for writing objectives and tests, record keeping, and analyzing test results.

Future studies could include comparing students' academic performance at Binford to students enrolled in schools of comparable size after sequentially-based objectives and criterion-referenced tests have been implemented for three to five years.

REFERENCES

REFERENCES

Boozer, R.F. & Lindvall, C.M. An investigation of selected procedures for the development and evaluation of heirarchical curriculum structures. New York, New York: American Educational Research Association, 1971. (ERIC Document Reproduction Service No. 049 287)

Bormuth, J.R. Feasibility of criterion-referenced testing and evaluation. Chicago, Illinois: University of Chicago, 1978. (ERIC Document Reproduction Service No. 159 225)

Bormuth, J.R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Cox, R.C. & Vargas, J.S. A comparison of item selection techniques for norm-referenced and criterion-referenced tests. Pittsburgh, Pennsylvania: Pittsburgh University, 1966. (ERIC Document Reproduction Service No. 010 517)

Crambert, A.C. Estimation of validity for criterion referenced tests. New York, New York: American Educational Research Association, 1977. (ERIC Document Reproduction Service No. 151 418)

Florida Department of Public Instruction. A guide to 1976-1977 statewide assessment results. Tallahassee, Florida: Student Assessment Section, Department of Instruction, 1976-77. (ERIC Document Reproduction Service No. 137 406)

Hennes, J.D. Accountability in colorado. New York, New York: American Education Research Association, 1977. (ERIC Document Reproduction Service No. 139 835)

Martuza, V.R. Applying norm-referenced and criterion-referenced measurement in education. Boston: Allyn & Bacon, 1977.

Petrosky, A.R. The 3rd national assessment of reading and literature versus norm- and criterion-referenced testing. Houston, Texas: International Reading Association, 1978. (ERIC Document Reproduction Service No. 159 599)

Skager, R.W. The great criterion-referenced test myth. Los Angeles, California: California University, 1978. (ERIC Document Reproduction Service No. 160 650)